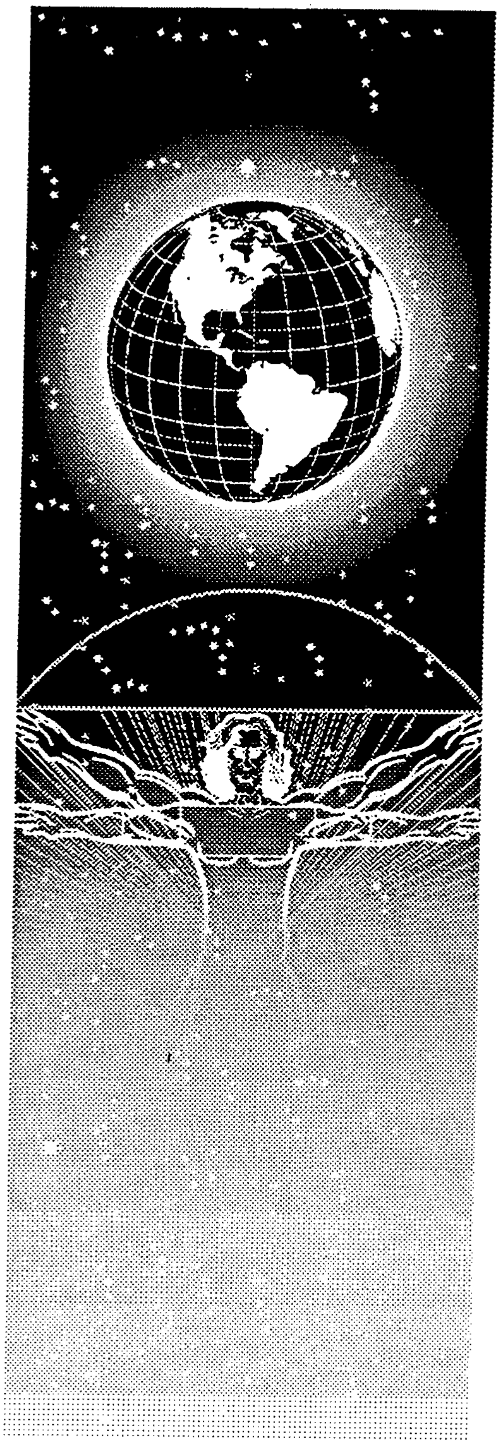
AL/HR-TP-1996-0017



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

THE ROLE OF OBSERVATIONAL LEARNING IN AUTOMATED INSTRUCTION OF COMPLEX TASKS

Barry P. Goettl
Cathy Connolly-Gomez

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

Galaxy Scientific Corporation
1880 Carswell Ave., Bldg. 9020
Lackland AFB, Texas 78236-5507

March 1998

AIR FORCE MATERIEL COMMAND
ARMSTRONG LABORATORY
HUMAN RESOURCES DIRECTORATE
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

BARRY P. GOETTL
Project Scientist

R. BRUCE GOULD
Technical Director
Cognition and Performance Division

WILLIAM E. ALLEY
Chief, Cognition and Performance Division

Please notify this office, AL/HRPP, 7909 Lindbergh Drive, Brooks AFB TX 78235-5352, if your address changes, or if you no longer want to receive our technical reports. You may write or call the STINFO office at DSN 240-3877 (or 240-2295) or commercial (210) 536-3877 (or 536-2295).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1998	Interim - July 1992 to September 1994

4. TITLE AND SUBTITLE

Role of Observational Learning in Automated Instruction of Complex Tasks

5. FUNDING NUMBERS

C - F41622-92-D-006
PE - 61102F
PR - 2313
TA - T2
WU - 18

6. AUTHOR(S)

Barry P. Goettl, Ph.D.
Cathy Connolly-Gomez

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Galaxy Scientific Corporation
1880 Carswell Avenue
Lackland AFB, TX 78236-5507

8. PERFORMING ORGANIZATION

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Technical Training Research Division
Brooks AFB, TX 78235-5352

10. SPONSORING/MONITORING

AL/HR-TP-1996-0017

11. SUPPLEMENTARY NOTES

Paper presented at the Human Factors and Ergonomics Society 39th Annual Meeting, October 12, 1995, San Diego, CA

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This study tested the prediction that observational learning will be more effective for motor tasks having substantial cognitive demands than for those that do not. Subjects were divided into three treatment groups: performers, observers, and no-observe controls. In Phase I, subjects were trained on a computer-based flight task requiring relatively little cognitive demands. In Phase II, subjects were trained on a different flight task that had significant cognitive and strategic demands. In Phase I, performers were superior to both observers and controls; the observers did not differ significantly from the controls. In Phase II, observation showed a beneficial effect for females. The female observers performed as well as the female performers. The results of this study suggest that observational learning benefits tasks with significant cognitive components more than tasks that are primarily psychomotor. Implications for computer-based training are discussed.

14. SUBJECT TERMS

Computer-Based Training, Gender Differences, Observational Leaarning, Skill Acquisition

15. NUMBER OF PAGES
14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Table of Contents

PREFACE

This report describes one of several experiments conducted in the TRAIN Cooperative Laboratory from October 1994 to March 1995. Funds for this research were provided by the U.S. Air Force Office of Scientific Research and the Armstrong Laboratory TRAIN Project, AL/HRTI, Brooks AFB, TX, Dr. Wes Regian, Director. A special thanks to Galaxy Scientific Corporation for data collection.

SUMMARY

This study tested the prediction that observational learning will be more effective for motor tasks having substantial cognitive demands than for those that do not. Subjects were divided into three treatment groups: performers, observers, and no-observe controls. In Phase I, subjects were trained on a computer-based flight task requiring relatively little cognitive demands. In Phase II, subjects were trained on a different flight task that had significant cognitive and strategic demands. In Phase I, performers were superior to both observers and controls; the observers did not differ significantly from the controls. In Phase II, observation showed a beneficial effect for females. The female observers performed as well as the female performers. The results of this study suggest that observational learning benefits tasks with significant cognitive components more than tasks that are primarily psychomotor. Implications for computer-based training are discussed.

THE ROLE OF OBSERVATIONAL LEARNING IN AUTOMATED INSTRUCTION OF COMPLEX TASKS[1]

INTRODUCTION

Computer-Based Training (CBT) systems have historically been geared toward training individuals rather than groups. Computer-Aided Instruction (CAI) systems are individualized in the sense that they are self paced and may incorporate branching routines for additional individualization of instruction. Intelligent Tutoring Systems (ITS), which are among the most intelligent CAI systems, epitomize individually tailored instruction through student modeling and remediation. While many ITSs have fared well on empirical evaluation (for a review see Shute and Regian, 1990), some question their cost effectiveness (Clancey, 1992). Budget constraints in many applied settings may limit the feasibility of providing every student with a computer. More importantly, a pedagogy designed for individuals may ignore important social factors, such as observational learning, that may be very beneficial in learning certain kinds of tasks. These considerations have lead some researchers to explore automated instructional pedagogies that are based on small groups rather than individuals (Shebilske and Regian, 1992).

Shebilske and his colleagues have begun to explore the role of observational learning in computer-based training. Shebilske, Regian, Arthur, and Jordan (1992), examined a dyadic Active Interlocked Modeling (AIM) protocol, in which each trainee alternatively controls half of a complex task (i.e., Space Fortress) while observing another subject performing the other half. This protocol allows each subject to learn critical task components by observing the actions and reactions of their partner and then putting that knowledge to practice when roles are reversed. Shebilske, Jordan, Arthur, and Regian (1993) expanded the dyadic protocol to a tetradic protocol by adding two passive observation roles. Subjects rotated through all four roles. The results revealed that four trainees could learn as well as one with one fourth the "hands on" practice and one fourth the trainer time and resources.

The goal of the present experiment was to explore the conditions under which observational learning is an effective pedagogy for computer-based flight simulator tasks. Bandura's multiprocess theory of observational learning (Bandura, 1986) provides a basis for predicting when and how observational learning may be used in computer-based training. In this theory observational learning is determined by four processes: attention, retention, behavioral production, and motivation. First subjects must be able to attend and extract the critical features of the task. Then this information must be transformed into appropriate cognitive representations that function as an internal model of the task. Next, behavioral production processes are needed to translate the symbolic representation into appropriate actions. During this process, actions and their consequences are compared to the internal model and behaviors are modified as a result of the comparisons. Finally, motivation processes influence the acquisition of the internal model by governing attention and retention processes. Motivation also influences performance of the action by governing the behavioral production processes.

This theory predicts that if the spatial and temporal features of the task can be easily extracted and coded, there will be little need for overt practice (Carroll and Bandura, 1985). However, if the features are subtle or intricate it will be difficult to translate the internal model into behaviors. Indeed, Martens, Burwitz, & Zuckerman (1976) showed that intricate motor responses can not easily be translated from cognitive representations into actions. In other words, observational learning will have more impact on tasks whose features can be easily extracted and symbolically coded than on tasks whose features are subtle. In support of this hypothesis, Bandura and Jeffery (1973) showed that responses were recalled better when they were symbolically coded than when they were not. Moreover, they showed that for coded responses, rehearsal of the symbolic code resulted in better performance than rehearsal of the motor response itself.

The present experiment tests the hypothesis that observational learning will have more benefit for tasks that can be symbolically coded than for tasks that cannot. We predict that observational learning will have more impact on flight simulator tasks that require substantial cognitive demands than simulator tasks that primarily require perceptual-motor responses. Thus, in this study there were two phases: one devoted to a flight task that is primarily perceptual-motor, and one devoted to a more cognitive-strategic flight task.

METHODS

Subjects

One hundred and two subjects (66 males and 36 females) completed the study. All subjects were recruited by local temporary employment agencies and were paid about $5.00 per hour for their participation. Subjects ranged in age from 18 to 30 years of age and had a high school diploma or Graduate Equivalency Degree (GED), but had not completed a four-year college degree. Subjects were screened to eliminate those who had previous experience with the simulator or reported spending more than 20 hours per week playing video games. Subjects were randomly assigned to three groups of 22 males and 12 females.

Equipment

The computer-based flight simulator, called Phoenix, was designed to train Heads-Up-Display (HUD) reading and basic flight skills. The Phoenix display provides an out-of-cockpit view of a simulated world. The HUD shows airspeed, heading, and altitude, as well as a climb/dive ladder for pitch and roll. The data were collected in the TRAIN CoLab at Lackland AFB, Texas. This laboratory contains 30 Compaq DeskPro 486/33L computers with NEC/Multisync VGA monitors and CH Products Flight Sticks.

Tasks

Slalom task. The slalom task required subjects to "fly" the simulator though "gates" in the sky. Subjects had to maneuver the simulator horizontally and vertically to fly through the gates. Four different courses (2 easy and 2 difficult) were used. Trials lasted 3 minutes and subjects were instructed to fly through as many gates as possible while minimizing misses. Speed (i.e., number of gates made) and accuracy (i.e., percent gates made) were measured and provided as feedback between trials.

Strike task. The strike task, required subjects to navigate the simulator through the simulated environment and shoot three targets. Trials ended when all targets were shot or after five minutes. Subjects were given feedback at the end of each trial. Feedback scores were computed based on the number of targets shot, total time (in seconds), and number of missiles fired [score = (100,000 * targets shot) / (time * missiles)]. In a transfer test, a more difficult version of the task was used. In this task subjects had to shoot four targets instead of three.

Procedure

In Phase 1 subjects were divided into three groups: performers, observers, and no-observe controls. The performers and observers were paired with each other. During training, the performers performed the slalom task 100% of the time while the observers watched. Subjects were not allowed to talk to each other. Meanwhile, the no-observe controls played a computer-based card game. There were two practice sessions separated by a test session. A second and a third test session occurred after the second practice session. During all three test sessions, all subjects were tested on four three-minute trials on the slalom task. Phase 1 ended with a "remediation" session in which the observers and controls practiced the slalom task while the performers played the computer-based card game. This remediation period was designed to reduce any group differences that occurred in Phase 1 and make the groups more comparable in total time on task. This remediation period was followed by two test sessions, and an 18-hour retention test session.

Phase 2 of the study was identical to Phase 1 except that the Strike task was used instead of the slalom task. In addition, to control for residual effects, 1/2 the performers in Phase 1 became observers in phase 2, and 1/2 the observers in Phase 1 became performers in phase 2. There were two practice sessions separated a test session and followed by a second test session. All test sessions in phase 2 consisted of three trials of the easy version and three trials of the difficult version of the strike task.

RESULTS

Phase 1

Slalom task performance. The Slalom task speed scores (i.e., average number of gates made) for the three groups during in Phase 1 of the study are shown in Figure 1. As expected, performers showed a large practice effect during the first two practice blocks. Across the first three test blocks (shown as T1, T2, and T3 in Figure 1), performers showed better overall performance ($M = 10.79$) than observers ($M = 5.81$) and controls ($M = 5.33$). The average speed scores on these tests were submitted to a 3 (training condition) x 2 (gender) x 3 (test) mixed factors ANCOVA with pretest scores serving as a covariate and tests serving as a repeated measures factor. This analysis showed significant main effects of training condition (F (2, 95) = 19.96, $p < .001$) and test (Wilks' exact F (2, 95) = 13.75, $p < .001$). No other main effects or interactions were significant. Planned contrasts on the training condition factor indicated that: (1) the performers were superior to the controls ($t = 5.96$, $p < .001$), and (2) there was no difference between the observers and the controls ($t = 1.14$, $p > .10$).

Efficacy of Remediation. During the remediation sessions only the observers and the control group received practice on the slalom task. The goal of remediation was to reduce any group differences for Phase 2 of the study. To determine whether the remediation sessions were successful in reducing group differences, performance on two immediate posttests and two delayed posttests (shown as T4 to T7 in Figure 1) were submitted to a 3 (training condition) x 2 (gender) x 2 (test: immediate vs. delayed) x 2 (trial) mixed factors ANCOVA with pretest scores as a covariate. The test and trial factors were repeated measures factors. This analysis showed only a significant main effect of trial (F (1, 96) = 4.82, $p = .031$) and a test x trial interaction (F (1, 96) = 14.67, $p < .001$). The main effect of group was not statistically significant (F (2, 95) = 1.46, $p = .237$), indicating that remediation was effective in reducing group differences.

Phase 2

Overall, performers showed higher average scores on the strike task than observers and controls (M = 488, 272, and 168 for performers, observers, and controls respectively). However, performers also had greater variance than subjects in the other two groups (SD = 377, 233, and 212 for performers, observers, and controls, respectively). Using Brown

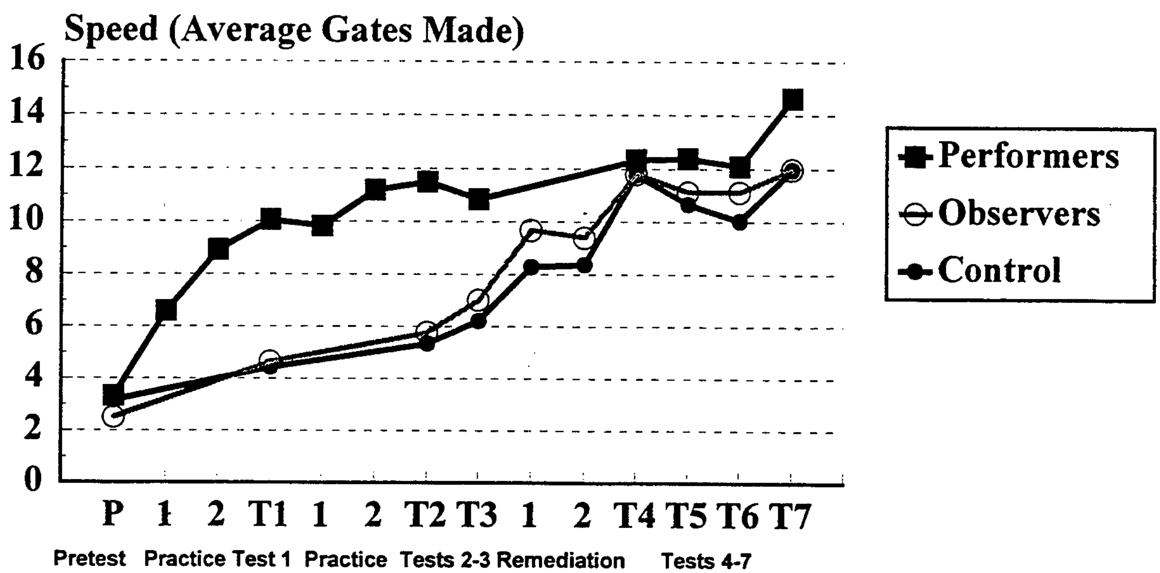


Figure 1. Average number of gates made on slalom task during pretest, acquisition, remediation, and posttest sessions for the three training groups.

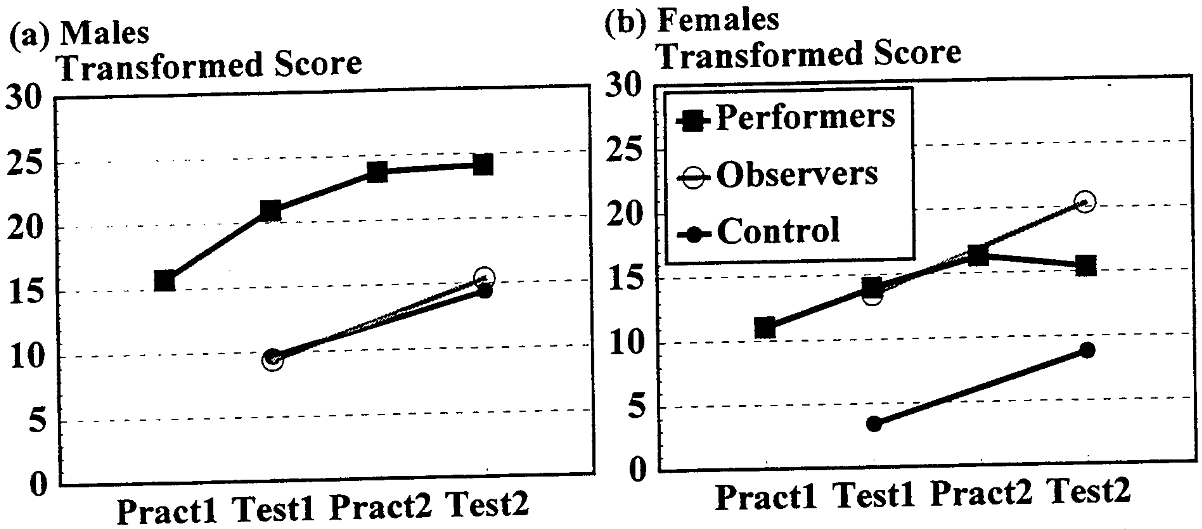


Figure 2. Transformed Strike task scores (square root of raw score) during practice and test blocks for males and females in the three training groups.

and Forsythe's (1974) test for differences among variances, we concluded that the variances between training groups were not homogeneous (F (2, 96) = 8.90, $p < .001$). Consequently, we performed a square root transformation on raw strike task scores to reduce group differences in variance. The transformed scores for practice and test sessions for males and females in the three groups are shown in Figure 2. Using Brown and Forsythe's test on the transformed scores, we concluded that group variances were homogeneous (F (2, 96) = 3.02, $p = .07$). All subsequent analyses were computed on the transformed scores.

Performers (M = 19.76) scored higher than observers (M = 13.96) and controls (M = 9.93). Transformed scores from the two test sessions were submitted to a 3 (training condition) x 2 (gender) x 2 (test) mixed factors ANOVA with test being the only repeated measure. This analysis revealed significant main effects of training condition (F (2, 96) = 11.85, $p < .001$) and test (F (1, 96) = 47.99, $p < .001$). In addition, the analysis showed reliable interactions between training condition and gender (F (2, 96) = 5.72, $p = .004$) and between training condition and test (F (2, 96) = 3.41, $p = .037$). The interaction between training condition and test indicated that the effect of training condition was larger for test 1 (F (2, 96) = 16.04, $p < .001$) than for test 2 (F (2, 96) = 6.83, $p = .002$).

To explore the training condition x gender interaction, we computed the main effect of training condition for males and females separately. Males showed a simple main effect of training condition (F (2, 63) = 12.00, $p < .001$) and planned contrasts revealed that: (1) the performers performed better than the controls ($t = 4.31$, $p < .001$) and (2) there was no difference between the observers and the controls ($t = 0.14$, $p = .89$). Females also showed a simple main effect of training condition (F (2, 96) = 7.71, $p = .002$), but planned comparisons showed a different pattern than the males: (1) the performers performed significantly better than the observers ($t = 2.96$, $p = .006$), and (2) the observers performed better than the controls ($t = 3.71$, $p < .001$).

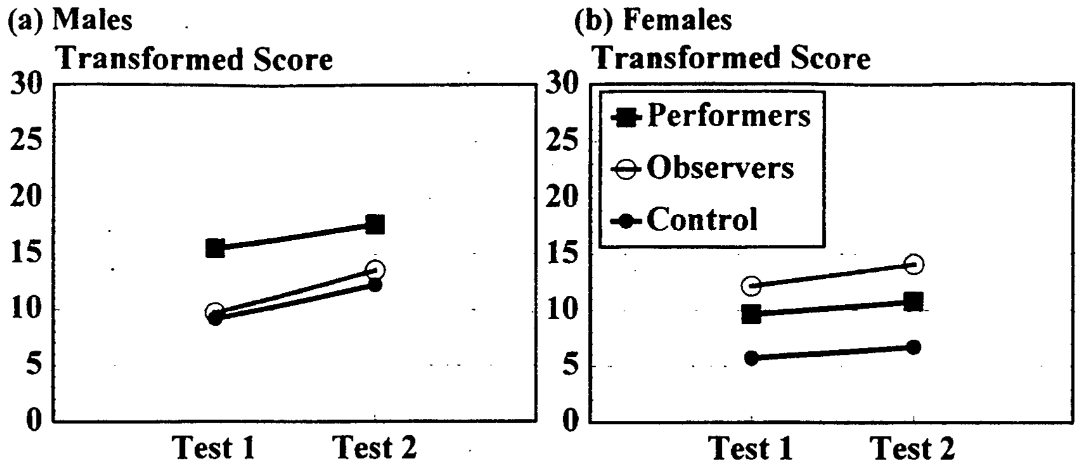


Figure 3. Transformed Strike task scores (square root of raw score) on the transfer task for males and females in the three training groups.

Data from the transfer strike task is shown in Figure 3. Again strike task scores were transformed to square roots. Analysis indicated that variances between groups on the transformed scores were homogeneous (F (2, 96) = 1.44, p = .24). Transformed scores were submitted to a 3 (training condition) x 2 (gender) x 2 (test) mixed factors ANOVA. This analysis revealed significant main effects of training condition (F (2, 96) = 4.99, p = .009), gender (F (1, 96) = 5.57, p = .020), and test (F (1, 96) = 15.05, p < .001). In addition there was a significant training condition x gender interaction (F (2, 96) = 3.12, p = .049). To further examine this interaction, the main effect of training condition was computed for males and females separately. For males, the main effect of training condition was significant (F (2, 63) = 4.42, p = .016) but planned contrasts indicated that the performers scored significantly higher than the controls (t = 2.76, p = .007) while the observers did not (t = 0.43, p = .67). For females the main effect of training condition was significant (F (2, 33) = 5.16, p = .011) and planned contrasts indicated that observers (t = 3.20, p = .003), but not performers (t = 1.84, p = .074), scored higher than the controls.

DISCUSSION

The data indicates that observation had no effect on the psychomotor task (i.e., slalom task). Observers' performance did not differ from the control subjects' performance. In fact, observers did not show much improvement until the remediation phase, wherein they nearly doubled their scores. However, even then, observers and control improved at the same rate. In contrast, observation did appear to benefit performance on the more cognitive task (i.e., strike task). However, only females showed the effect. Specifically, females observers performed better than female controls on both versions of task (i.e., easy and difficult), but female performers differed from female controls on only the easy version of the task.

The failure to find a significant observational effect for the slalom task is consistent with the hypothesis that observation should have no effect on tasks that are difficult to represent symbolically. The slalom task is primarily a visual-motor task. It involves rapid generation of motor responses based on visual and kinesthetic feedback. The goal can be easily verbalized (i.e., fly through as many gates as possible) but symbolic representation of the behaviors is difficult and involves subtle and intricate responses that require overt rehearsal. While this conclusion is based on the failure to reject a null hypothesis, it is strengthened by the fact that controls improved as fast as observers during the remediation phase. If observers successfully generated a symbolic representation of the task from their observations, it did not appear to help them during the behavioral production phase of observational learning. In short, observation of the slalom task was of little practical benefit whatsoever.

The differential effectiveness of observation for males and females on the strike task was unexpected and is difficult to explain. One possible explanation is gender differences in skill level. On average, males performed better on the strike task than females. Because observers were randomly assigned to performers, it could be argued that males observed players with higher skills less often than did females. Thus, on average, males may have had less opportunity than females to benefit from observation. This conclusion is less than satisfying for two reasons. First, male observers did not perform better than male control subjects. Some proportion of the male observers were paired with subjects of higher skill and should have benefited from observation, yet there was no apparent benefit. Second, female observers' scores were higher, though non-significantly higher, than male observers' scores.

A related argument pertains to the nature of skill acquisition. Several theories of skill acquisition posit that initial stages of skill acquisition focus on declarative knowledge (Anderson, 1983). The emphasis on symbolic representations in Bandura's theory suggests that observational learning may be more effective during the declarative knowledge stage of skill acquisition. In context of the current data, perhaps males can by-pass the declarative knowledge stage more quickly than females because of higher aptitude for the task (Goettl, Yadrick, Gomez, Regian, and Shebilske, in press). This would explain the gender difference in observational learning. However, the present data provide little support for or against this hypothesis. Another reason for the differential effect of observation for males and females may be related to how males and females perform the task. Perhaps females attended to the verbal-symbolic processes in the task while males attended to the nonverbal processes. In other words, perhaps male observers do not generate or utilize a symbolic representation of the task but females do. However, without verbal protocols from subjects concerning what they learned from observation or how they performed the task, the current data can neither confirm nor disconfirm this hypothesis.

CONCLUSIONS

Observational learning has been shown to be an effective technique for motor tasks that can be symbolically coded (Carroll and Bandura, 1985; Bandura and Jeffery, 1973) as well as motor tasks that have significant cognitive components (Shebilske and Regian, 1992; Shebilske et al., 1993). The present data are consistent with these previous findings in that observational learning was more beneficial for a task with significant cognitive processing demands than one that primarily requires psychomotor skills. The major aspect of psychomotor tasks may be the acquisition of coordination between visual and proprioceptive cues. If so, such tasks may fall outside the domain of observational learning.

The present results provide additional evidence that observational learning may be an effective tool to use in conjunction with automated instruction. However, its use may be limited to certain individuals and certain tasks. Identification of these limits will aid in the application of group pedagogy into CBT.

ACKNOWLEDGMENTS

Funds for this research were provided by the U.S. Air Force Office of Scientific Research and the Armstrong Laboratory TRAIN Project, AL/HRTI, Brooks AFB, TX.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Bandura, A. (1986). *Social Foundation of thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A., and Jeffrey, R. W. (1973). Role of symbolic coding and rehearsal processes in observational learning. *Journal of Personality and Social Psychology, 26,* 122-130.

Brown, M. B., and Forsythe, A. B. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association, 69,* 364-367.

Carroll, W. R., and Bandura, A. (1985). Role of timing of visual monitoring and motor rehearsal on observational learning of action patterns. *Journal of Motor Behavior, 17,* 269-281.

Clancey, W. J. (1992). The frame of reference problem in the design of intelligent machines. In K. VanLehn (Ed.), *Architectures for intelligence: The Twenty-Second Carnegie-Mellon Symposium on Cognition.* Hillsdale, NJ: Erlbaum.

Goettl, B. P., Yadrick, R. M., Gomez, C. C., Regian, J. W., and Shebilske, W. L. (in press). Alternating task modules in isochronal distributed training of complex tasks. *Human Factors.*

Martens, R., Burwitz, L., and Zuckerman, J. (1976). Modeling effects on motor performance. *Research Quarterly, 47,* 277-291.

Shebilske, W. L., Jordan, J. A., Arthur, W. Jr., and Regian, J. W. (1993). Combining a multiple emphasis on components protocol with small group protocols for training complex tasks. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society* (1216-1220). Santa Monica, CA: The Human Factors and Ergonomics Society.

Shebilske, W. L., and Regian, J. W. (1992). Video games, training, and investigating complex skills. *Proceedings of the 36th Annual Meeting of the Human Factors and Ergonomics Society* (1296-1300). Santa Monica, CA: The Human Factors and Ergonomics Society.

Shebilske, W. L., Regian, J. W., Arthur, W. Jr., and Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors, 34,* 369-374.

Shute, V. J., and Regian, J. W. (1990, June). *Rose garden promises of intelligent tutoring systems: Blossom or thorn?* Paper presented at the Space Operations Automation and Robotics Conference, Albuquerque, NM.